U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
SYSTEMS DEVELOPMENT OFFICE
TECHNIQUES DEVELOPMENT LABORATORY

TDL OFFICE NOTE 82-2

RECENT EXPERIMENTS IN THE USE OF MODEL OUTPUT STATISTICS FOR FORECASTING SNOW AMOUNTS

Joseph R. Bocchieri

February 1982

1. INTRODUCTION

A Model Output Statistics (MOS) system for forecasting heavy snow has been operational within the National Weather Service since October 1977 (Bocchieri, 1979a); National Weather Service, 1978). Heavy snow is defined as a fall of ≥4 inches during a 12-h period at a station. In MOS (Glahn and Lowry, 1972), a statistical relationship is determined between the forecast output of a numerical prediction model (or models) and observed occurrences of a particular weather element. For the operational heavy snow system, we used output from both the Limited-area Fine Mesh (LFM) model (National Weather Service, 1971; Gerrity, 1977) and a finer mesh version of the LFM, called LFM-II (National Weather Service, 1977; Newell and Deaven, 1981), to develop prediction equations that give the conditional probability of heavy snow for 12-24 h periods after 0000 GMT and 1200 GMT. An estimate of the unconditional probability of heavy snow and a categorical forecast are also provided. The unconditional probability is derived by taking the product of the conditional probability, the probability of precipitation (PoP) (Lowry and Glahn, 1976; National Weather Service, 1980), and the conditional probability of frozen precipitation (PoF) (Bocchieri, 1979b).

We've recently derived a set of experimental snow amount forecast equations with the goal of developing a new operational system which would provide forecasts for other snow amount categories, not just heavy snow, and other forecast projections other than 12-24 hours. In the process of deriving these equations, we performed a number of experiments. In one experiment, we determined the effect of removing LFM boundary-layer type predictors from the forecast equations, since future operational numerical models at the National Meteorological Center may not include a boundary layer in the form presently used in the LFM. The results indicate that forecast equations that didn't include boundary-layer predictors were only slightly less accurate than equations that included boundary-layer predictors. In another experiment, we determined the optimum number of predictors to include in the forecast equations and found that 12 predictors was about right.

Conditional probability of heavy snow forecasts made by the experimental equations were then compared to similar forecasts made by the operational system; the results indicate that there was little difference in the accuracy of the two systems. This was encouraging because the experimental equations didn't include boundary-layer type predictors.

After performing these experiments for the 12-24 h projection, we developed conditional probability equations for several snow amount categories for the 12-18, 18-24, and 24-36 h projections from 0000 GMT. We then compared two methods for estimating the unconditional probability of occurrence of the snow amount categories, given the conditional probabilities from the experimental system. One method, called PRODUCT, consisted simply of multiplying the conditional probabilities for each snow amount category by the PoP for the

Table 1 were included in both binary and continuous form. The use of binary predictors helps to account for non-linear relationships which may exist between the predictors and predictands. A good description of the screening procedure can be found in Glahn and Lowry (1972).

a. Preliminary Experiments

In the process of developing the experimental PoSA(S) equations, we did a number of experiments. We first determined the effect of removing LFM boundary-layer predictors from the PoSA(S) equations. Two PoSA(S) forecast systems were developed with data combined from 195 conterminous United States stations and seven winter seasons, September through April, 1972-73 through 1978-79. One system included boundary-layer predictors, among others, and the other system didn't. The REEP technique was used in conjunction with the potential predictors in Table 1 to develop 12-term equations for each system. The two PoSA(S) systems forecast the probability of ≥ 4 inches of snow, given that snow occurs, in the 12-24 h period after 0000 GMT.

We did a comparative verification between the two PoSA(S) systems with independent data from the winter season of 1979-80. The results show that the Brier score (Brier, 1950) for the PoSA(S) system which included boundary-layer predictors was about 1% better than the Brier score of the system that didn't have these predictors. We judged that this improvement was not large enough to include boundary-layer predictors in further development of PoSA(S) equations, since these fields may not be available from future numerical models run at the National Meteorological Center.

In another experiment, we determined the optimum number of predictors to include in PoSA(S) equations. First, PoSA(S) equations were derived for the 12-24 h period after 0000 GMT for each of several geographic regions. The developmental sample consisted of data from 195 stations and eight winter seasons (1972-73 through 1979-80). The regions were determined in the following manner. We derived PoSA(S) equations for the ≥ 1 and ≥ 4 inch categories for the 12-24 h period from both 0000 GMT and 1200 GMT with data combined from 195 stations for the developmental sample--the so-called generalized operator approach. We then evaluated the equations to obtain forecasts for each station on the developmental sample. A statistic called the relative probability bias was computed for each station and for each snow amount category by:

$$\text{Relative Probability Bias} = \frac{\overline{\text{PoSA(S)}} - \text{RF}}{\text{RF}}, \qquad (1)$$

where $\overline{\text{PoSA(S)}}$ is the average conditional probability forecast for a particular snow amount category for each station and RF is the relative frequency of that category for each station for the developmental sample. We subjectively determined the regions shown in Fig. 1 by grouping stations having similar relative probability bias characteristics; other factors we considered were the density of stations and the climatic frequency of snow amount. We needed to make the regions large to insure that a sufficient number of snow amount cases would be available for equation development.

weather systems producing lesser snow amounts.

In Table 4, the predictor types are ranked as determined by the REEP screening procedure for the 12-24 h and 24-36 h projections from 0000 GMT. This list was determined by both frequency and order of selection; for the purpose of this ranking, all predictor prjections, smoothings, and binary limits were combined for each type of variable. The results indicate the LFM forecasts of P AMT and MEAN R HUM were the most important predictors for both projections. Various parameters derived at 850-mb, such as 850 DIV, 850 VORT, and 850 M CONV, ranked next in importance at 12-24 hours, and 850 U and 850 DIV ranked next in importance at 24-36 hours. Other investigators such as Browne and Younkin (1970), Brandes and Spar (1971), and Spiegler and Fisher (1971) also found the 850-mb level to be useful for snow amount prodiction. Variables at 700-mb, such as 700 U and 700 VV followed the 850-mb parameters for 12-24 hours. For the 24-36 h projection, the 500 VORT ranked next followed by 700 U. Goree and Younkin (1966) and Weber (1978) also found various parameters at 500-mb to be useful. 10-5 THICK ADV ranked relatively high for the 12-24 h projection; Younkin (1968) found strong warm advection to be associated with heavy snow in the western United States.

3. EXPERIMENTS IN ESTIMATING THE UNCONDITIONAL PROBABILITY OF SNOW AMOUNT

We experimented with two methods for obtaining unconditional probability of snow amount, PoSA, forecasts for the 12-24 h period after 0000 GMT. One method, called PRODUCT, consists simply of multiplying the conditional probability forecast, PoSA(S), for each snow amount category, by the PoP for the corresponding 12-h period and the average PoF ($\overline{\text{PoF}}$) for the same 12-h period. This is expressed mathematically[1], for instance for the $\geq$1 inch category, by:

$$\text{PoSA}(\geq 1 \text{ inch}) = \text{PoSA(S)}(\geq 1 \text{ inch}) \times \text{PoP} \times \overline{\text{PoF}}. \qquad (2)$$

This method is essentially the one used in the operational system to obtain estimates of the unconditional probability of heavy snow.

The other method, called MOSSQR, involved derivation of REEP equations to predict PoSA. Potential predictors for these equations were PoSA(S), PoP, and PoF forecasts, and various products thereof. The developmental sample for this derivation consisted of the months October through March, 1972-73 through 1979-80. We first developed generalized-operator PoSA equations with data combined from 192 stations for the $\geq$1, $\geq$2, $\geq$3, $\geq$4, and $\geq$6 inch categories. We then attempted to determine regions in a manner similar to that used for the PoSA(S) equations in Section 2. That is, we computed the relative probability bias for each station by Eq. (1) using the developmental sample and then tried to combine stations with similar bias characteristics to determine the regions

---

[1] In a true mathematical sense, this method doesn't give the unconditional probability of snow amount for 12-h periods in which mixed precipitation occurs, since PoSA(S) equations were developed with pure snow cases only.

a category, say ≥ 1 inch of snow, is a value that if exceeded by a probability forecast for that category would result in a categorical forecast of ≥ 1 inch. If the threshold value is not exceeded, the categorical forecast would be <1 inch.

Threshold probabilities were computed with the use of an empirical iterative technique. On each iteration, threat scores were computed for categorical forecasts made by comparing probability forecasts against threshold probabilities. For the initial iteration, a first guess threshold probability was provided. For subsequent iteration, the threshold probability was incremented by a preselected value. Threshold values were chosen so that the categorical bias was ≤ 1.30, even if the maximum threat score was associated with a bias >1.30.

As as example, assume that the probability forecasts for the 12-24 h period for the categories ≥ 1, ≥ 2, ≥ 3, ≥ 4, and ≥ 6 inches are 0.40, 0.30, 0.20, 0.15, and 0.05, respectively. Furthermore, assume that the threshold probabilities that maximize the threat score for each of the five categories are 0.38, 0.32, 0.23, 0.16, and 0.08. The procedure starts at the category ≥ 1 inch and compares the probability forecast (0.40) with the threshold value for that category (0.38). Since the threshold value is exceeded, at least 1 inch is predicted. The next step is to proceed to the category ≥ 2 inch and see if the probability forecast (0.30) exceeds the threshold value (0.32) for that category. Since it doesn't, the procedure is terminated and the forecast amount is 1 to 2 inches. Of course, if the probability forecasts had exceeded the threshold values for all categories, the forecast amount would have been ≥ 6 inches. Similarly, if the probability forecast was less than the threshold value for the category ≥ 1 inch, the forecast amount would have been <1 inch.

After determining the threshold probability values, we produced and verified categorical snow amount forecasts for both the developmental and independent data samples, which were the same as those used in Section 3. Table 7 shows the threat score and bias for the 12-24, 12-18, 18-24, and 24-36 h forecast projections from 0000 GMT. For the two 12-h periods, we verified five categories of snow amount, ≥ 1, ≥ 2, ≥ 3, ≥ 4, and ≥ 6 inches; for the two 6-h periods, we verified three categories, ≥ 1, ≥ 2, and ≥ 3 inches. The results indicate the following:

1. For the developmental sample, the biases were generally between 1.00 and 1.30. The bias was forced to be such when the threshold probability values were derived. For these bias values, the threat scores were between .20 and .30 for the ≥ 1, >2, and ≥ 3 inch categories for the 12-24 and 24-36 h projections and for the ≥ 1 and ≥ 2 inch categories for the two 6-h periods. The threat scores were generally between .15 and .20 for the remaining categories for each projection.

2. For the independent sample, both the threat score and bias deteriorated as compared to the developmental sample. The deterioration was worse for the 18-24 h and 24-36 h periods than for the 12-18 h and 12-24 h periods. Also, the deterioration was generally worse

accurate as the operational forecasts, in spite of the fact that the experimental equations didn't include boundary-layer type predictors while the operational equations did. We therefore proceeded with the development of experimental PoSA(S) equations for other projections.

PoSA(S) equations were developed for the 12-18 h, 18-24 h, and 24-36 h projections from 0000 GMT. An analysis of the predictor types chosen by the REEP screening procedure for the 12-24 h and 24-36 h projections showed that LFM forecasts of precipitation amount and surface to 500-mb mean relative humidity were the most important predictors for both projections. Various parameters derived at the 850-mb level, such as divergence, vorticity, moisture convergence, and wind components ranked next in importance. Similar variables at the 700-mb and 500-mb levels followed.

We also expermented with two methods for obtaining <u>unconditional</u> probability of snow amount, PoSA, forecasts for the 12-24 h period after 0000 GMT. One method, called PRODUCT, consists simply of multiplying PoSA(S) forecasts for each snow amount category by the PoP for the corresponding 12-h period and the average conditional probability of frozen precipitation, $\overline{\mathrm{PoF}}$, for the same 12-h period. This is quite similar to the method used in the present operational heavy snow system. Another method, called MOSSQR, involved derivation of regression equations to p edict PoSA. Potential predictors for these equations were PoSA(S), PoP, and $\overline{\mathrm{PoF}}$ forecasts, and various products thereof. A comparative verification between PRODUCT and MOSSQR on independent data indicates there was little difference in Brier scores for the two systems. Hence, we decided to use the PRODUCT method to estimate PoSA, since it will be easier to put into operation.

To help us determine whether the experimental snow amount forecast system would be suitable for possible operational implementation, we verified categorical snow amount forecasts on independent data. First, we used the PRODUCT method to estimate PoSA for each projection, each region, and for both the developmental and independent data samples. The PoSA forecasts were then transformed into categorical snow amount forecasts; to do this, we derived threshold probability values for each snow amount category, each region, and each forecast projection using the developmental sample. The threshold values were chosen such that the threat score was maximized while restricting the bias to ≤ 1.30. We then computed the threat score and bias for categorical snow amount forecasts for the independent data sample. The results indicate that both the threat score and bias deteriorated as compared to these scores for the developmental sample. The deterioration was worse for the 18-24 h and 24-36 h periods than for the 12-18 h and 12-24 h periods. Also, the deterioration in the scores was generally worse for the higher snow amount categories compared to the lower categories for each forecast period. We also compared the threat score and bias for the ≥ 4 inch category for the experimental system to the same scores computed for the operational heavy snow system. The results indicated that the two systems were of comparable accuracy.

Lowry, D. A., and H. R. Glahn, 1976: An operational probability of precipitation--PEATMOS PoP. Mon. Wea. Rev., 104, 221-231.

Miller, R. G., 1964: Regression estimation of event probabilities. Tech. Rep. No. 1, Contract CWB-10704, The Travelers Research Center, Inc. Hartford, Conn., 153 pp. [NTIS AD 602037].

National Weather Service, 1971: The limited-area fine mesh (LFM) model. NWS Tech. Proc. Bull., No. 67, NOAA, U.S. Department of Commerce, 11 pp.

_____, 1977: High resolusion LFM (LFM-II). NWS Tech. Proc. Bull., No. 206, NOAA, U.S. Department of Commerce, 6 pp.

_____, 1978: The use of model output statistics for predicting the probability of heavy snow. NWS Tech. Proc. Bull., No. 246, NOAA, U.S. Department of Commerce, 10 pp.

_____, 1980: The use of model output statistics for predicting probability of precipitation. NWS Tech. Proc. Bull. No. 289, NOAA, U.S. Department of Commerce, 13 pp.

Newell and Deaven, 1981: The LFM-II model--1980. NOAA Tech. Memo. NWS NMC-66, NOAA, U.S. Department of Commerce, 20 pp.

Spiegler, D. B., and G. F. Fisher, 1971: A snowfall prediction method for the atlantic seaboard. Mon. Wea. Rev., 99, 311-325.

Weber, E. M., 1978: Major midwestern snowstorms. Preprints Conference on Weather Forecasting and Analysis and Aviation Meteorology, American Meteorological Society, Silver Spring, 30-37.

Younkin, R. J., 1968: Circulation patterns associated with heavy snowfall over the western United States. Mon. Wea. Rev., 96, 851-853.

Zurndorfer, E. A., and R. J. Bermowitz, 1976: Determination of an optimum number of predictors for probability of precipitation amount forecasting. TDL Office Note 76-17, National Weather Service, NOAA, U.S. Department of Commerce, 7 pp.

0000 GMT.

Number of Predictors	Brier Scores				
	Snow Amount Category (inches)				
	≥1	≥2	≥3	≥4	≥6
1	.406	.232	.133	.072	.023
2	.395	.232	.133	.073	.023
4	.389	.232	.133	.073	.022
6	.388	.233	.132	.071	.022
8	.387	.230	.132	.071	.023
10	.386	.229	.131	.071	.023
12	.385	.230	.132	.071	.023
16	.383	.229	.131	.072	.023
20	.384	.229	.131	.071	.023

Table 3. The reduction of variance for REEP conditional probability of snow amount equations for the 12-24 h projection from 0000 GMT for each of the five regions shown in Fig. 1. The developmental data were from 8 winter seasons (1972-73 through 1979-80). The relative frequency (%) of each snow amount category is also shown in parentheses.

Region	Total Reduction of Variance (%)					Total Number of Snow Cases
	Snow Amount Category (inches)					
	≥1	≥2	≥3	≥4	≥6	
1	10.9 (37.3)	9.9 (16.9)	7.3 (7.4)	8.6 (3.1)	7.4 (0.9)	1307
2	19.2 (49.3)	15.8 (27.9)	14.5 (17.8)	11.6 (10.3)	6.8 (3.4)	1100
3	17.8 (36.7)	17.9 (17.9)	17.9 (9.2)	18.3 (4.6)	13.2 (1.4)	4170
4	18.0 (39.4)	24.9 (19.6)	20.6 (11.1)	18.9 (6.4)	18.2 (2.3)	3785
5	22.8 (36.4)	24.1 (17.8)	22.5 (9.8)	20.8 (5.9)	17.3 (2.5)	3799

stations for October through March, 1972-73 through 1979-80. The additional reduction of variance given by each predictor is shown. The number of cases for each snow amount category is given in parentheses.

Predictors	Additional Reduction of Variance (%)				
	Snow Amount Category (inches)				
	$\geq$1	$\geq$2	$\geq$3	$\geq$4	$\geq$6
	(6599)	(3170)	(1679)	(899)	(307)
PoSA(S) ($\geq$1 in) x PoP x $\overline{\text{PoF}}$	27.4	19.8	13.7	9.3	4.5
PoSA(S) ($\geq$6 in) x PoP x $\overline{\text{PoF}}$	0.1	1.5	2.8	4.5	7.2
PoSA(S) ($\geq$2 in) x PoP x $\overline{\text{Pof}}$ $\leq$70%	0.0	0.0	0.1	0.6	1.4
PoSA(S) ($\geq$3 in) x PoP x $\overline{\text{PoF}}$	0.0	0.5	0.8	0.5	0.0
PoSA(S) ($\geq$4 in) x PoP x $\overline{\text{PoF}}$ $\leq$50%	0.0	0.0	0.0	0.0	0.5
PoSA(S) ($\geq$2 in) x PoP x $\overline{\text{PoF}}$ $\leq$50%	0.0	0.0	0.1	0.3	0.1
PoSA(S) ($\geq$6 in) x PoP x $\overline{\text{PoF}}$ $\leq$4.5%	0.0	0.0	0.0	0.0	0.2
PoSA(S) ($\geq$2 in) x PoP x $\overline{\text{PoF}}$ $\leq$18%	0.1	0.2	0.1	0.1	0.1
PoSA(S) ($\geq$4 in) x PoP x $\overline{\text{PoF}}$	0.0	0.0	0.0	0.2	0.0
Climatic Frequency $\geq$6 in	0.0	0.0	0.0	0.1	0.1
PoSA(S) ($\geq$6 in) x PoP x $\overline{\text{PoF}}$ $\leq$1.5%	0.0	0.0	0.0	0.0	0.1
PoSA(S) ($\geq$2 in) x PoP x $\overline{\text{PoF}}$ $\leq$80%	0.0	0.0	0.0	0.0	0.1

combined from 192 stations.

Forecast Projection from 0000 GMT	Verification Score	≥1 D	≥1 I	≥2 D	≥2 I	≥3 D	≥3 I	≥4 D	≥4 I	≥6 D	≥6 I
		Snow Amount Category (inches)									
12-24 h	Threat score	.30	.28	.25	.19	.21	.14	.19	.15	.18	.12
	Bias	1.08	.83	1.10	.65	1.07	.61	1.09	.64	1.03	1.00
	Number of snow cases	6599	629	3174	298	1671	150	902	83	310	27
12-18 h	Threat score	.27	.26	.21	.16	.17	.10	--	--	--	--
	Bias	1.10	.77	1.13	.71	1.02	.56	--	--	--	--
	Number of snow cases	3470	306	1353	126	584	55	--	--	--	--
18-24 h	Threat score	.26	.19	.21	.11	.18	.12	--	--	--	--
	Bias	1.22	.63	1.05	.39	1.20	.40	--	--	--	--
	Number of snow cases	2980	305	1212	119	528	47	--	--	--	--
24-36 h	Threat score	.28	.22	.22	.13	.20	.05	.17	.00	.12	.00
	Bias	1.29	1.09	1.14	.75	1.19	.57	1.16	.25	1.11	.14
	Number of snow cases	4387	655	2082	278	1108	133	564	69	185	21

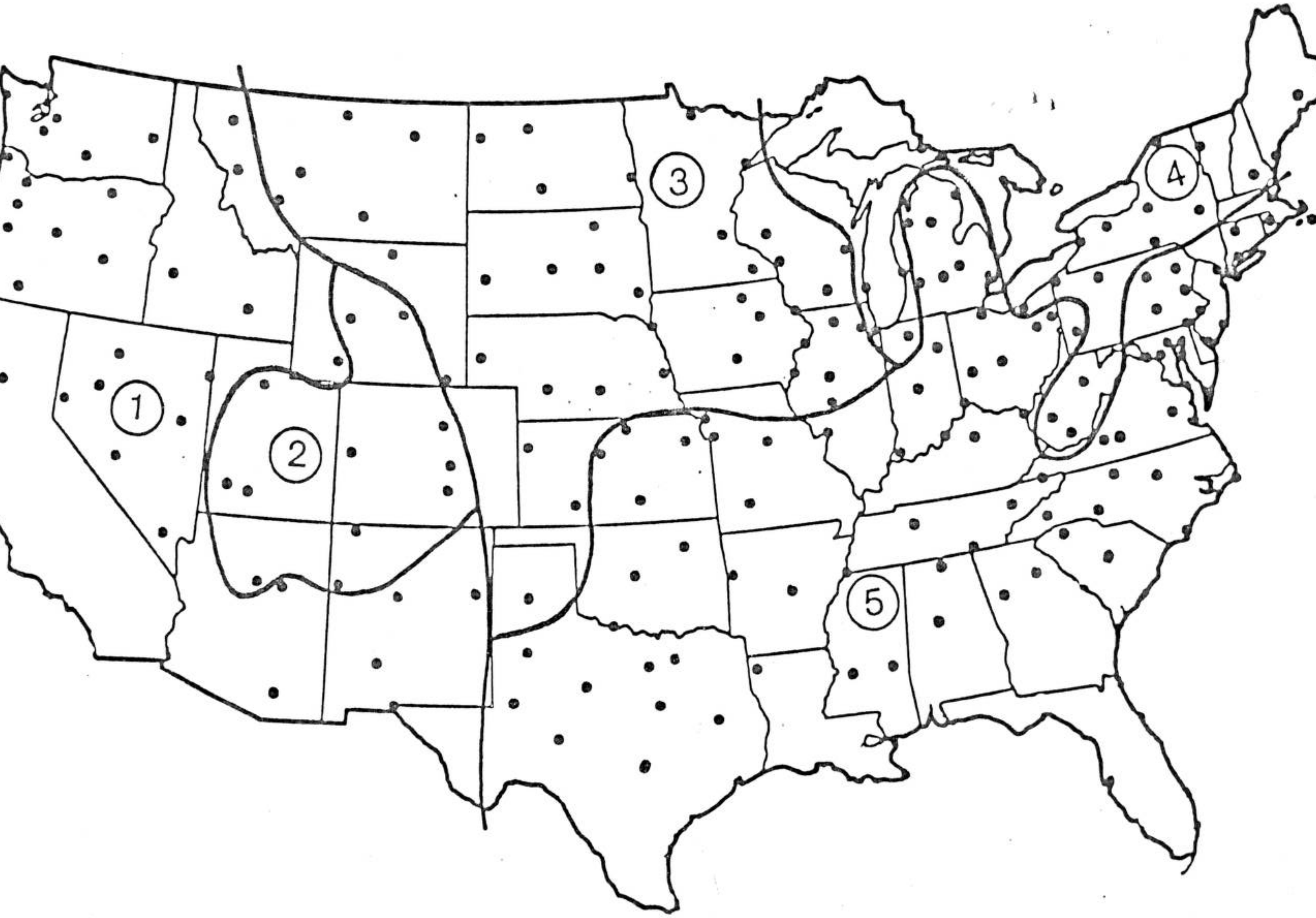

Figure 1. The five regions used in the development of the experimental PoSA(S) equations. The dots show the stations for which snow amount data were available in the developmental archive.